Visions for language

Paul Mc Kevitt
Department of Computer Science
Regent Court, 211 Portobello Street
University of Sheffield
GB- S1 4DP, Sheffield
England, EU.
E-mail: p.mckevitt@dcs.shef.ac.uk

Abstract

One of the most interesting developments in the field of Artificial Intelligence (AI) has been that the field has split into a number of subfields with little or no communication between them. For example, two of the most important subfields, that of natural language processing and that of vision processing have been tackled on their own and only recently have we seen a major move towards their integration. We discuss here the reason why we believe this has happened, representations for grounding language and vision, lexicons, and some current projects on mathematical, applied and interface issues in integration. The 1990s bode well for systems with integrated visions for language.

1 Introduction

Although there has been much progress in developing theories, models and systems in the areas of Natural Language Processing (NLP) and Vision Processing (VP) (see Partridge 1991, Rich and Knight 1991) there has been little progress on integrating these two subareas of Artificial Intelligence (AI). Although in the beginning the general aim of the field was to build integrated language and vision systems, few were done, and two subfields quickly arose. It is not clear why there has not already been much activity in integrating NLP and VP. Is it because of the long-time reductionist trend in science up until the recent emphasis on chaos theory, non-linear systems, and emergent behaviour? Or, is it because the people who have tended to work on NLP tend to be in other Departments, or of a different ilk, to those who have worked on VP? There has been a recent trend towards the integration of NLP and VP and other forms of perception such as speech (see Dennett, 1991; Mc Kevitt 1994a, 1994b, 1994c, 1994d, 1994e, Mc Kevitt and Guo 1994 and Wilks and Okada 1994). Dennett (1991, p. 57-58) says "Surely a major source of the widespread skepticism about "machine understanding" of natural language is that such systems almost never avail themselves of anything like a visual workspace in which to parse or analyze the input. If they did, the sense that they were actually understanding what they processed would be greatly heightened (whether or not it would still be, as some insist, an illusion). As it is, if a computer says, "I see what you mean" in response to input, there is a strong temptation to dismiss the assertion as an obvious fraud."

Heretofore social trends in science in general have been towards reductionism. Pure reductionism argues that the social nature of experimentation is irrelevant to scientific outcome (see Popper 1972) where the interactions between scientists should have no effect upon their results. In contrast, the effort on integration required here will certainly involve social interaction between researchers in each field which might not have occurred otherwise (see Kuhn, 1962). What matters to scientific progress is not the conducting of experiments per se, but rather the determination of which experiments are worth conducting. In such contexts, 'worth' is clearly a sociological, as opposed to a scientific, matter. Our hope is for a reversal of the unfortunate reductionist influence, through the re-unification of currently disparate strands of enquiry. Already, science has started to move in this direction (see, for example, Gleick, 1987; Langton, 1989 and Rowe and Mc Kevitt 1991).

2 Background

People are able to combine the processing of language and vision with apparent ease. In particular, people can use words to describe a picture and can reproduce a picture from a language description. Moreover, people can exhibit this kind of behaviour over a very wide range of input pictures and language descriptions. Even more impressive is the fact that people can look at images and describe not just the image itself but a set of abstract emotions evoked by it. Although there are theories of how we process vision and language there are few theories about how such processing is integrated. There have been large debates in Psychology and Philosophy with respect to the degree with which people store knowledge as propositions or pictures (see Pylyshyn 1973, Kosslyn and Pomerantz 1977).

There are at least two advantages of linking the processing of natural languages to the processing of visual scenes. First, investigations into the nature of human cognition may benefit. Such investigations are being conducted in the fields of Psychology, Cognitive Science, and Philosophy. Computer implementations of integrated VP and NLP can shed light on how people do it. Second, there are advantages for real-world applications. The combination of two powerful technologies promises new applications: automatic production of text from images; automatic production of images from text; and the automatic interpretation of images with text. The theoretical and practical advantages of linking natural language and vision processing have also been described in Wahlster (1988).

Early work for synthesising simple text from images was conducted by Waltz (1975) who produced an algorithm capable of labelling edges and corners in images of polyhedra. The labelling scheme obeys a constraint minimisation criterion so that only sets of consistent labellings are used. The system can be expected to become 'confused' when presented with an image where two mutually exclusive but self-consistent labellings are possible. This is important because in this respect the program can be regarded as perceiving an illusion such as what humans see in the Necker cube. However, the system seemed to be incapable of any higher-order text descriptions. For example, it did not produce natural language statements such as "There is a cube in the picture."

A number of natural language systems for the description of image sequences have been developed (see Herzog and Retz-Schmidt 1990, Neumann and Novak 1986). These systems can verbalize the behaviour of human agents in image sequences about football and describe the spatio-temporal properties of the behaviour observed. Retz-Schmidt (1991) and Retz-Schmidt and Tetzlaff (1991) describe an approach which yields plan hypotheses about intentional entities from spatio-temporal information about agents. The results can be verbalized in natural language. The system called REPLAI-II takes observations from image sequences as input. Moving objects from two-dimensional image sequences have been extracted by a vision system (see Herzog et al. 1989) and spatio-temporal entities (spatial relations and events) have been recognised by an event-recognition system. A focussing process selects interesting agents to be concentrated on during a plan-recognition process. Plan recognition provides a basis for intention recognition and plan-failure analysis. Each recognised intentional entity is described in natural language. A system called SOCCER (see André et al. 1988, Herzog et al. 1989) verbalizes real-world image sequences of soccer games in natural language and REPLAI-II extends the range of capabilities of SOCCER. Here, NLP is used more for annotation through text generation whereas we are interested in analysis.

Maaβ et al. (1993) describe a system, called *Vitra Guide*, that generates multimodal route descriptions for computer assisted vehicle navigation. Information is presented in natural language, maps and perspective views. Three classes of spatial relations are described for natural language references: (1) topological relations (e.g. in, near), (2) directional relations (e.g. left, right) and (3) path relations (e.g. along, past). The output for all presentation modes relies on one common three-dimensional model of the domain. Again, Vitra emphasises annotation through generation of text, rather than analysis, and the vision module considers interrogation of a database of digitized road and city maps rather than vision analysis.

Some of the most recent engineering work in NLP is focusing on the exciting idea of incorporating NLP techniques with speech, touchscreen, video and mouse to provide advanced multimedia interfaces (see Maybury 1992). Examples of such work are found in the ALFresco system which is a multimedia interface providing information on Italian Frescoes (see Carenini et al. 1992 and Stock 1991), the WIP system that provides information on assembling, using, and maintaining physical devices like an expresso machine or a lawnmower (see André and Rist 1992 and Wahlster et al. 1992), and a multimedia interface which identifies objects and conveys route plans from a knowledge-based cartographic information system (see Maybury 1991). We can now move on to investigate a representation for integrated language and vision processing.

3 Language and vision processing

Traditionally, language processing has considered the phenomenon of mapping natural language into semantic representations of the objects, actions and states that that language describes (see e.g. Allen 1987, Gazdar and Mellish 1989, Mc Kevitt 1992, Mc Kevitt et al. 1992a, Schank 1972, 1973, 1975 and Wilks 1973, 1975a, 1975b, 1975c). A question that has arisen time and time again is that of the nature of these representations. Due to the reductionist nature of scientific enquiry, the answer to this question has always been given in a reduced propositional form, usually containing base meanings for words called *semantic primitives* (see Wilks 1977), which is itself a form of language. This leads to the inevitable circularity that (the representation is a language (is a language (is a language which represents the world))). Such a ludicrous situation clearly calls for the interpretation of language-based representations in terms of other language-independent representations. One such representation is vision.

In situ, vision processing has considered the possibility of mapping visual scenes into objects, actions and states. Since the visual scene is described not in terms of yet more visual descriptions, but rather in terms of propositions, the above circularity of language to language representation is avoided. However, this does not, of itself, avoid the problem - it merely delays it. For now, we see that vision is explained in terms of language, which is either circular, or else re-expressed in terms of vision!

3.1 Representation

Where does leave us? It seems that traditional NLP involves the mapping of language into propositional structures, while VP also involves the mapping of visual scenes into symbolic structures. In each case, the final representation is language-based. But then that language, if we follow what we did already, could be represented in another language and so on. An obvious question suggests itself - can these circularities be avoided by the resolution of adequate 'base cases'? In other words, is there a uniform representation in the form of propositions, or even pictures, to which all other representations can be reduced? It would certainly be more economic if there was just one representation for perceptions, whether this be in the form of pictures or propositions, or perhaps some other formalism capable of subsuming these two as distinct instances of a more general scheme. We suggest that, since each formalism can clearly be mapped to the other, both propositional and visual representations are useful for this purpose, and it doesn't really matter which is used as an internal representation. Probably, both will be useful, although the tradition in the field of AI has been to use propositions which we believe has caused an unfair bias towards language.

3.2 From language to vision...

If we take some utterance, such as "the black cat", it is easy to imagine its pictorial representation. For a given agent, the pictorial representation of the black cat will generally be similar to that of any other agent of at least the same culture. Obviously, however, for more personal subject matter, for example my aspirations, there is every possibility that different scenes may be produced. However, if agents are considered to be sharing their individual experiences with each other then the scenes will be understood to be valid representations of utterances.

We differ with Dennett (1991), therefore, who claims that there is no visual representation of the meaning of concepts such as yesterday. Dennett (1991, p. 57) says, "Uncles, unlike clowns and firemen, don't look different in any characteristic way that can be visually represented, and yesterday's don't look like anything at all." Indeed, visual representations are easy to discover. For example, an icon-based system might use a picture of a diary to remind users of the next day's events, thereby representing tomorrow. Or perhaps a cartoon sequence might represent yesterday by showing the world running rapidly backwards where the re-emergence of daylight indicates that the previous day has been reached, at which point the regress slows and eventually terminates. Such pictorial representations of the passage of time, and our mind's eye location within them, are commonplace in the film industry. Iconic and animated representations of things we normally understand in language are given in Beardon (1994) and Narayanan et al. (1994).

3.3 ...and back again

Similarly, given any particular picture, for example an abstract evocation of the dawn, two agents would not necessarily produce equivalent propositional descriptions of it. However, it does not follow that the methods they are using to translate between representational mechanisms are incompatible. For example, two people may produce completely conflicting descriptions of a scene, and yet still be willing to accept an argument as to why the given description is valid in the context of the other agent.

Where does that leave us? Well, it seems that language can be translated into vision and vision into language and that although there may be discrepancies in these translations these may be due to the agent of translation. And where are language and vision grounded? Maybe a useful way to answer that question might be to take a look at one of our age old standards of language...

3.4 Lexicons

In any integrated natural language and vision processing system there will probably be some move towards an integrated lexicon. Today's dictionaries are sorely lacking in information that people have had in their heads for years. If one thinks of a "lion", a "zebra", "parties", "love", "hate", "sex", "loud", "bang", "greasy", "furry", "running", "jumping", "swooning", "ice cream", etc. then one has a picture in one's head of these objects, emotions, sounds, feelings and actions or some situation when they occurred in past personal history or in a film. Such pictures and sounds, and their manifestation in the symbols of written language itself were a major part of the emphasis of the writings of Joyce (1922, 1939) and others. For example, Joyce (1950) uses letters in English to produce the sounds of the waves as they come rushing towards the seashore on Dollymount Strand.

Today's dictionaries such as Longman's Dictionary of Contemporary English (LDOCE) (see Procter 1978), Collins COBUILD (see Sinclair, 1987) and Webster's, whether in text form or in electronic form, do not contain much pictorial information; they typically encode words in symbolic natural language form with symbolic natural language descriptions. Encyclopedias do contain pictures but they do not contain definitions of words, rather knowledge about words, and specifically objects in the world. It is not clear to us why dictionaries have had this bias towards symbolic natural language but it certainly seems very strange behaviour.

There has been much work in AI and in NLP on defining dictionaries for use in large intelligent systems. Examples are the Machine Tractable Dictionaries (MTDs) produced from the Machine Readable Dictionaries (MRDs) such as LDOCE (see Guo 1992, Guthrie et al. 1991). In fact the idea of processing MRDs to obtain lexicons for NLP systems has become one of the largest research areas in NLP. There has also been work on encoding large encyclopedias for AI systems (see Lenat and Guha 1989).

One of the suggested solutions to problems of NLP over the years have been to reduce word and sentence representations to primitives (see Wilks 1977). Schank defined 14 of such primitives for Conceptual Dependency (see Schank 1972, 1973, 1975) and Wilks had some 80 in his Preference Semantics system (see Wilks 1973, 1975a, 1975b, 1975c). However, all this reductionist work had difficulties because of at least the following two reasons: (1) *circularity*: some words are defined in terms of primitives but those primitives are defined in terms of the original words; (2) *grounding*: how are the primitives grounded in the world? i.e. what gives them their meaning? We discuss the circularity problem here and the grounding problem is handled in the next section.

How come in dictionaries you look up a word like 'gorse' and find that the definition of that word involves 'furze' and when you look up 'furze' its definition uses 'gorse'? In LDOCE, the primitive for 'disease' is defined to be 'disorder' or 'illness' and these in turn are defined as 'disease'. This has been a problem for dictionaries for years. Katz (1972) claims that linguistic primitives play the role that neutrinos play in science. Wilks (1994) points out that primitives in natural language do not have any obvious visual analogues and that no definition of primitives is necessary because they are explained by the procedural role they play in language as a whole.

We argue in Mc Kevitt and Guo (1994) that defining vocabularies and primitives can be defined in terms of spatial or pictorial representations to obtain meanings. So, for example, taking a primitive for the concept of abstract transfer (called ATRANS by Schank) we can have a picture showing two agents with an object being transferred between the two. This picture could be animated as demonstrated by Beardon (1994) and Narayanan et al. (1994) and could be shown to a user on demand. Furthermore, a definition of the changes in spatial relations with respect to ATRANS could be represented. For example, this would detail the the object, instrument, trajectory and duration of transfer as defined in the perceptual semantics of Chakavarthy (1994). There are also full blown multi-modal lexicons under development. Young (1983) and Wright and Young (1990) describe a knowledge representation called Cognitive Modalities (CM) for neural networks which is a cognitive, non-verbal representation of information.

Hence, one can define a word by using a definition that uses other words but also spatial and visual structures. These structures would give partial definitions of words so that there would only be at most partial circularity in definitions. Such pictorial information is missing from today's dictionaries. The ability to develop and learn new words such as metaphors is to a large extent based on spatial and pictorial mappings. Our systems of the future will need to be able to apply algorithms for such mappings to existing dictionaries to derive new ones. And, of course, Wittgenstein (1953, p. 42) pointed out already that "It is only in the normal cases that the use of a word is clearly prescribed."

3.5 Symbol grounding

Harnad (1990, 1993) has brought the grounding problem further and said that symbolic processing systems have, in general a problem with grounding their symbols and that this problem can only be freed up by using other perceptual sources such as visual input. This is his answer to Searle's Chinese Room[1] problem where Searle argues that a machine cannot *understand* the symbols it represents but just passes them around with no feeling for their meaning (see Searle 1980, 1984, 1990). In some sense what Searle is arguing is that the computer behaves like a hypertext system does, encoding text and being able to manipulate and move it around but having no sense for its meaning. Jackson and Sharkey (1994) argue that connectionist architectures are necessary for grounding perceptual systems or at least that such grounding will be easier with such architectures. Harnad's solution to the Chinese Room problem has also been suggested by Marconi (1994) and Meini and Paternoster (1994).

One feature that is perhaps less obvious is the need for analog mechanisms for representation, as highlighted in recent exchanges (see Harnad 1993). The reductionist paradigm has led many to believe, unquestioning, that digital approximations to an analog world are good enough, in the sense that any property of the world can be 'unpacked' and approximated arbitrarily closely by digital representations. Stannett (1990) suggests, however, that machine-models based on the analog paradigm are strictly more powerful than their digital counterparts, in the sense that analog machinery is capable of performing tasks that are provably impossible in a digital world. Accordingly, if we are to represent the analog mechanisms inherent for example in neurological systems (see, for example, MacLennan 1990), it will be necessary to adopt radically new mechanisms for the integrated representations involved.

Hence, we are settling on a solution to grounding where language is grounded in vision and vision in language and that they feed on each other for interpretation, explanation and understanding. Some concepts are easier expressed in language representations, and some in pictures, and the fields of AI and computation have tended to rely too much on the former. Let's move on to look at some projects we are conducting in language and vision in order to get a feel for understanding them more.

4 Some projects in language and vision

We now describe three projects in language and vision. The first project focusses on theoretical representations for language and vision integration, the second focusses on an application, and the third on interfaces.

4.1 Algebraic semantics

Morgan and Mc Kevitt (1994a) have begun developing an algebraic semantics of spatial relations using topology theory in mathematics and corresponding matrix operations. Algebraic structure is exhibited by a collection of operations that act on objects transforming one member of a set to another. We investigate *groups* and how operations on them combine members of the set to produce other members with the properties of inverse, identity, associativity and closure. Examples of groups are rotations of objects in two dimensions, or rotations and reflections in two dimensions. Also, investigated are *rings* and how they represent infinite sets of objects. Finally, we look at *fields* and how they are used for enlarging and reducing objects. In Morgan and Mc Kevitt

[1] Searle asked us to imagine a Chinese Room where a person who cannot understand Chinese is locked in the room and has the task of using an English rule book for manipulating Chinese symbols. Then, to an outside observer, the person appears to be able to understand Chinese just as a computer program which manipulates symbols could appear to do so (see Searle 1984, 32-33).

(1994b) we discuss how Euclidian geometry can be applied to objects in an environment and we look in particular at reflections, rotations, translations, and glide reflections. The algebraic semantics is currently applied to spreadsheet display with spatial relations formulated in terms of matrix operations and could also be used for giving a spatial definition of primitives such as ATRANS developed by Schank. We are currently investigating this possibility.

4.2 Interpretation of angiograms

We are considering a model of a system that processes X-ray projections (angiograms) and their associated medical reports (see Hall and Mc Kevitt 1994). Each of these input data relate to blood vessel structures (vasculature) and arteriovenous malformations (AVM) within the human body. AVMs are congenital abnormalities on the vasculature. These AVMs are dangerous because if they hemorrhage the results can be fatal. The primary clinical motivation for acquiring angiograms is to plan radiation therapy of the AVMs. Our goal is to reconstruct the original vasculature and AVMs in three-dimensional space. The angiograms to be processed by the VP module of the integrated system are gray-level digital images produced by a radiographer. Also, it is standard for radiologists to prepare medical reports on the angiograms. Typically, the medical reports are spoken into a dictating machine and afterwards transcribed into written form by a secretary. Hence, the input the NLP module of the integrated system is written text. Within this application domain we believe that an integrated system will perform better from the synergy of NLP and VP rather than each working individually.

Often there will only be two angiograms per patient that are perspective projections. One is taken from the front of the body called the posterior-anterior view. The other is from the side and is called a lateral view. Together these are called biplane angiograms and are separated by an angle of about ninety degrees and a small lateral mis-alignment. A contrast agent makes blood vessels visible in X-ray angiograms. There are a host of problems with reconstruction from digital angiograms such as warping of individual projections, loss of geometric relationship between biplane pairs, and phase difference of contrast agent between biplane angiogram pairs. These are fully described in Hall (1993).

Standard methods available for reconstruction from images by reprojection are called filtered back-projection and algebraic reconstruction tomography (ART). Both of these fail for angiograms because each requires a large number of projections whereas we typically have less than ten projections per patient. However, reconstruction can be achieved if we can decide which correspondence points in each angiogram of the biplanar pair were generated by the same point in the three-dimensions being x-rayed. This is the correspondence problem that has been a ubiquitous problem within vision research (see Marr 1982, Mayhew and Frisby 1981).

There are two distinct ways to solve correspondence. First, points in each angiogram may be labelled. Then points with the same label correspond by definition. Alternatively, if we know the approximate whereabouts of points in three-dimensional space then a solution to correspondence will follow. Both of these methods require some kind of a priori information and we choose the latter method for its elegance and generality. We propose using a three-dimensional model that is an information repository about the physical vasculature. This model is constructed by combining many graphs of individual vasculatures into a total graph. Using this technique the model is capable of learning new vascular forms (see Hall et al. 1993b).

Reconstruction of the AVMs must proceed along very different lines. This is because the AVMs are essentially random distributions of tangled vessels. Consequently no a priori information regarding the three-dimensional form of AVMs can be used; that is each AVM is unique. Our VP module is capable of segmenting AVMs from their vascular context in angiograms. It can also reconstruct an approximation of the AVM volume (see Hall et al. 1993a). It must be approximate because of the low number of angiograms we have per patient.

Medical reports are produced by a radiologist who analyses the biplane angiograms for each patient. The reports are a description of a clinician's mental reconstruction from the biplane angiograms. The reports focus on the presence or absence of AVMs. They describe the location of AVMs in the vasculature by referring to the names of surrounding vessels and tissue such as brain matter. Also, the radiologist will often note unusual vascular formations. Angiograms usually depict or demonstrate the arterial system and not the venous system. Hence, the presence of veins will be noted in the reports. There will be a series of reports for each patient over the course of treatment. A possible problem with medical reports is that radiologists can make mistakes while analysing angiograms.

We are developing a computational model, in Prolog, for translating medical reports in English into meaning representations. We intend to implement that model in Quintus Prolog. The model is similar in spirit to that incorporated in the OSCON (Operating System CONsultant) system which has been implemented in Quintus Prolog and which answers English questions about computer operating systems (see Mc Kevitt 1986, 1991a, 1991b, Mc Kevitt and Wilks 1987, and Mc Kevitt et al. 1992a, 1992b, 1992c, 1992d).

The aim our work is to combine NL and VP modules into an integrated system. We expect a better performance to arise as a result of this combination. For example, in the medical domain just described we believe the subgraph isomorphism problem will benefit through the presence of more information. Also the VP module will act as a check on the accuracy of medical reports. The total graph model that acts as a knowledge representation is in another module and is a central resource to be used by the NLP and VP modules. These three modules will communicate with each other.

We believe that the way forward for developing general theories of language and vision processing is to focus on specific applications such as the medical domain. We hope that the integrated system will also shed light on how people process language with images and will help resolve some of the debate on how we represent knowledge about the linguistic and visual world.

4.3 A sensitive interface

One of the most important problems in human-computer interaction is that of maximising communication between the user and the computer. We claim that optimum communication will be facilitated when the computer can analyse the intentions of the computer user. We propose a philosophy for computer interface design where the computer analyses the intentions of users through verbal and nonverbal media. With respect to verbal media we have already mentioned a computer program called OSCON which can analyse users' intentions about computer operating systems. With respect to nonverbal media we argue that computers will be better able to analyse people intentions when recognising the media of facial expression, touch, and sound. A philosophy of interface design based on human-centeredness which brings the human and computer closer to analysing each others intentions, is argued for. As a test domain we have the IDIOMS[2] (Intelligent Decision-making In On-line Management Systems) project (see Gammack et al. 1989, 1991), which is applied to the domain of credit control. We argue that this approach will ensure that computers will become more understanding of their users and this will result in a more sensitive human-computer interface (see Mc Kevitt and Gammack 1994).

A theory of intention analysis (see Mc Kevitt 1991b) has been proposed as a model, in part, of the coherence of natural-language dialogue. A central principle of the theory is that coherence of natural-language dialogue can be modelled by analysing sequences of intention. The theory has been incorporated within a computational model in the OSCON system.

The computational model has the ability to analyse sequences of intention. The *analysis of intention* has at least two properties: (1) that it is possible to recognise intention, and (2) that it is possible to represent intention. The syntax, semantics and pragmatics of natural-language utterances can be used for intention recognition. Intention sequences in natural-language dialogue can be represented by what we call *intention graphs*. Intention graphs represent frequencies of occurrence of intention pairs in a given natural-language dialogue. An ordering of intentions based on *satisfaction* exists, and when used in conjunction with intention sequences, indicates the *local*[3] and *global* degree of expertise of a speaker in a dialogue.

The architecture of the OSCON system consists of six basic modules and two extension modules. There are at least two arguments for modularising any system: (1) it is much easier to update the system at any point, and (2) it is easier to map the system over to another domain. The six basic modules in OSCON are as follows: (1) ParseCon: natural-language syntactic grammar parser which detects query-type, (2) MeanCon: a natural-language semantic grammar (see Brown et al. 1975, and Burton 1976) which determines query meaning, (3) KnowCon: a knowledge representation, containing information on natural-language verbs, for understanding, (4) DataCon: a knowledge representation for containing information about operating system commands, (5) SolveCon: a solver for resolving query representations against knowledge base representations, and (6) GenCon: a natural-language generator for generating answers in English. These six modules are satisfactory if user queries are treated independently, or in a context-free manner. However, the following two extension modules are necessary for dialogue-modelling and user-modelling: (1) DialCon: a dialogue modelling component which uses an intention matrix to track intention sequences in a dialogue, and (2) UCon: a user-modeller which computes levels of user-satisfaction from the intention matrix and provides information for both context-sensitive and user-sensitive natural-language generation.

There are many more dimensions to communication than the merely verbal or linguistic we have just discussed. There is a whole realm of semiotics, which the computer in its current form does not begin to touch, and for users who prefer to operate through channels other than the superficially verbal, the computer is frustrating in its insensitivity. It is not our intention to get into a soporific debate about whether computers can have emotions, rather to consider what might be realistic in enabling them to have a more sensitive response.

First there is sensitivity to the expression of the user. Humans can usually tell from nonverbal cues if someone is tense, angry or impatient. This gives clues to formulate a response that is neither long winded nor patronizing, but considered and calming. This would be particularly useful in help or consultant systems. There is both an existing science and an ancient lore of how to read faces to detect characteristics (see Tao 1989). Some skeptical and empirical studies might indicate whether there is anything in this approach, and we have begun some cross-cultural work[4] in this direction. A computer that could detect visually the expression of its user begins to make inroads on a traditionally human quality, and introducing a seeing component into a computer also has implications for computer security.

Other cues to the mood of the user may be indicated by touch. For example, a user who is in a bad mood could indicate this by over-vigorous keypresses or bashing the mouse and this could be detected. Sensors can readily detect this sort of symptom, and use it in forming

[2] The IDIOMS project is being undertaken by the Bristol Transputer Centre at Bristol Polytechnic, The National Transputer Centre at Sheffield, Strand Ltd., and a well-known British high street retail bank.

[3] By *local* expertise we wish to stress the fact that sometimes experts can act as novices on areas of a domain which they do not know well.

[4] We are currently working with Dr. Wang in the Peoples Republic of China on this cross-cultural work. We have conducted initial experiments where subjects are asked to group 70 different faces represented by icons. Results show that several of the icons were recognised and grouped consistently by subjects. Items within such icons may become useful for icons of emotional expression.

a user model. Another cue which indicates user-intention is *sound.* User satisfaction levels such as irritation, hesitancy, and other personal characteristics can potentially be detected through voice interfaces. Again, such interfaces have implications for security.

Although perhaps fanciful, there is a serious point behind these ideas which has a definite message for design: typed interaction is a very superficial level of communication and limits the experience of interaction for many users. Although work in NLP is beginning to address intentions which underlie utterances, this is only one dimension of communication available to humans, and if computers are to be really user friendly, they have to become a lot more sensitive to their users' needs.

The IDIOMS project is one which uses techniques from AI to detect patterns in large databases, and to convey this information in a humanly understandable manner. Although supporting rule-based descriptions of expertise, IDIOMS represents a break from traditional expert system developments in several important regards. First, it uses a more flexible representation of knowledge, allowing the same knowledge base to conclude on numerous goal variables, and has straightforward extensibility and maintainability. Second, it allows the user to take the initiative and alter the thresholds and criteria for decision making. This *human centred design* coupled with an adaptive machine learning component allows many of the traditional problems of expert systems development to be overcome, such as knowledge acquisition, insensitivity to context, and inevitable obsolescence, while retaining the desirable features of heuristic processing, natural-language like communication, and reasoning transparency. The underlying hardware, a powerful parallel processing engine, allows fast management access to information without interfering with routine processing jobs, and empowers complex decision making.

The IDIOMS project has been applied to the domain of credit control. A typical problem for a bank manager or insurance underwriter involves looking at the entries on a loan application form and deciding whether it is a good or a bad risk. This is a complex function of predictors such as salary or outgoings, but may also be affected by indications such as phone ownership, and length of time at an address. It is our intention to incorporate techniques for non-verbal and verbal communication into the interface for the domain of credit control. We hope to incorporate techniques for intention analysis already existing in the OSCON system into the IDIOMS project.

It has been pointed out recently by Schank and Fano (1994) that in order to perform tasks in the world, understanding is a question of relating visual and linguistic input to the intentions derived from the task. They point out that expectations are a large part of understanding and say "We need to be able to reason about the things we can sense and the situations in which we will be able to detect them. Reminders to ourselves such as strings around fingers, notes on doors, alarm clocks, etc. all betray an underlying model of what we will notice (e.g. strings, sounds, notes) as well as the situations in which we will notice them (e.g. we are all likely to see our finger, pass through a door before leaving, and hear an alarm clock next to our beds."

We agree with Schank and Fano and believe that our own work in intention modelling can only be fulfilled by incorporating the analysis of visual scenes as well as symbolic natural language. In particular, our beliefs about people before they say anything at all are based on body language, clothes, looks, makeup, style and so on and work on modelling beliefs in language (see Ballim and Wilks 1990, 1991 and Wilks and Ballim 1987) will need to be augmented and integrated with work on determining beliefs from visual input. Indeed, work has already begun on determining intentions from vision and language (see Gapp and Maaβ 1994, Herzog and Wazinski 1994 and Maaβ 1994).

5 Conclusion

We have discussed the problem of integration of language and vision processing in AI and how the problem came about due to reductionist tendencies in this field and other areas of Science. Now we find an upsurge in work where people are attempting to build integrated systems again.

One of the most difficult problems in any AI system is the development of a dictionary or lexicon for containing definitions of meanings of the words for communication. We have pointed out that lexicons have been too biased towards language and need now to involve new structures with spatial definitions and iconic or animated representations. These structures would enable the reduction in circularity of definitions within existing dictionaries. The representations will also enable easier understanding of metaphors and derivation of new words from existing ones.

The symbol grounding problem was discussed and how the integration of language and vision could ground symbols in meanings that would help solve the Chinese Room problem. Symbols would be grounded in spatial or pictorial representations and that at least would reduce some of the grounding problem. Of course other sources such as sound, smell and touch would need to be added to give complete grounding.

Finally we discussed three projects for language and vision: (1) algebraic structures for an integrated semantics, (2) automatic interpretation of angiograms and (3) a sensitive interface. These projects and others will enable the fruitful integration of language and vision processing to continue to give us new visions for language.

6 References

Allen, James F. (1987) *Natural language understanding.* Benjamin/Cummings Series in Computer Science. Menlo Park, California: Benjamin/Cummings.

André, E., G. Herzog, and T. Rist (1988) On the simultaneous interpretation of real-world image sequences and their natural language description: the system SOCCER. In *Proceedings of the 8th European Conference on Artificial Intelligence*, 449-454, Munich, Germany.

André, Elisabeth and Thomas Rist (1992) The design of illustrated documents as a planning task. In *Intelligent multimedia interfaces*, M. Maybury (Ed.), 75-93. Menlo Park, CA: AAAI Press.

Ballim, Afzal and Yorick Wilks (1990) Stereotypical belief and dynamic agent modeling. In *Proceedings of the Second International Conference on User Modelling*, University of Hawaii at Manoa, Honolulu, Hawaii, USA.

Ballim, Afzal and Yorick Wilks (1991) *Artificial Believers.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Beardon, Colin (1994) Discourse structures in iconic communication. In *Artificial Intelligence Review, Special Volume on the Integration of Language and Vision Processing, Mc Kevitt, Paul (Ed.)*, Vol. 8, No. 3 (forthcoming).

Brown, John Seely, Richard R. Burton and Alan G. Bell (1975) SOPHIE: a step towards creating a reactive learning environment. In *International Journal of Man-Machine Studies*, 7, 675-696.

Burton, R. (1976) *Semantic grammar: An engineering technique for constructing natural-language understanding systems.* BBN Report No. 3453, Bolt, Beranek and Newman, Cambridge, MA, USA.

Carenini, G., F. Pianesi, M. Ponzi and O. Stock (1992) *Natural language generation and hypertext access.* IRST Technical Report 9201-06, Instituto Per La Scientifica E Tecnologica, Loc. Pant e Di Povo, I-138100 Trento, Italy, EU.

Chakavarthy, Anil (1994) Towards a perceptually-based semantics of action verbs. In *Proceedings of the Workshop on integration of natural language and vision processing*, Mc Kevitt, Paul (Ed.), Twelfth American National Conference on Artificial Intelligence (AAAI-94). Seattle, Washington, USA, August

Dennett, Daniel (1991) *Consciousness explained.* Harmondsworth: Penguin.

Gammack, J.G., S.A. Battle, and R.A. Stephens (1989) A knowledge acquisition and representation scheme for constraint based and parallel systems. In *Proceedings of the IEEE Conference on Systems Man and Cybernetics*, Hyatt Regency Hotel, Cambridge, Massachusetts, 1030-1035.

Gammack, J.G. T.C. Fogarty, S.A. Battle, and R.G. Miles (1991) Management decision support from large databases: the IDIOMS project. In *Proceedings of the Association of Modelling and Simulation Enterprises International Conference on Signals and Systems*, Institute of Transport, Warsaw, Poland, 213-219.

Gazdar, Gerald and Chris Mellish (1989) *Natural language processing in PROLOG: an introduction to computational linguistics.* Wokingham, United Kingdom: Addison-Wesley.

Gapp, Klaus-Peter and Maaβ, Wolfgang (1994) Spatial layout identification and incremental descriptions. In *Proceedings of the Workshop on integration of natural language and vision processing*, Mc Kevitt, Paul (Ed.), Twelfth American National Conference on Artificial Intelligence (AAAI-94). Seattle, Washington, USA, August

Gleick J. (1987) *Chaos - making a new science.* Harmondsworth, UK: Viking/Penguin.

Guo, Cheng-ming (1992) *Machine tractable dictionaries.* Norwood, N.J.: Ablex.

Guthrie, Joe, Louise Guthrie, Yorick Wilks, and H. Aidinejad (1991) Subject-Dependent Co-Occurrence and Word Sense Disambiguation. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, 146-152, Berkeley, California. Also Memoranda in Computer and Cognitive Science, MCCS-91-206, CRL, NMSU.

Hall, P.M. (1993) *Problems in reconstructing vasculature and lesions from angiograms.* Technical Report CS-93-05, Department of Computer Science, University of Sheffield, GB- S1 4DP, Sheffield, England, UK, EU.

Hall, P.M., John Brady, A.H. Watt, L. Walton, and U. Bergvall (1993a) *Segmenting and reconstructing vascular lesions from biplane angiogram projections.* Technical Report CS-93-01, Department of Computer Science, University of Sheffield, GB- S1 4DP, Sheffield, England, UK, EU.

Hall, P.M., A.H. Watt, J. McGregor, L. Walton, and U. Bergvall (1993b) *A model of a collection of vasculature.* Technical Report CS-93-03, Department of Computer Science, University of Sheffield, GB- S1 4DP, Sheffield, England, UK, EU.

Hall, Peter and Paul Mc Kevitt (1994) The application of integrated language and vision processing to reconstruction of angiograms. In *Artificial Intelligence Review, Special Volume on the Integration of Language and Vision Processing*, Vol. 8, No. 2 (forthcoming).

Harnad S. (1990) The symbol grounding problem. In *Physica D.*, 335-46.

Harnad S. (1993) Grounding Symbols in the Analog World with Neural Nets: A Hybrid Model. In *Think 2*, 12-20.

Herzog, G. and G. Retz-Schmidt (1990) Das System SOCCER: Simultane Interpretation und naturalichsprachliche Beschreibung zeitveranderlicher Szenen. In *Sport und Informatik*, J. Perl (Ed.), 95-119. Schorndorf: Hofmann.

Herzog, G., C.-K. Sung, E. Andre, W. Enkelmann, H.-H. Nagel, T. Rist, and W. Wahlster (1989) Incremental natural language description of dynamic imagery. In *Wissenbasierte Systeme. 3. Internationaler GI-Kongress*, C. Freksa and W. Brauer (Eds.), 153-162. Berlin: Springer-Verlag.

Herzog, Gerd and Peter Wazinski (1994) VIsual TRAnslator: Linking perceptions and natural language descriptions. In *Artificial Intelligence Review, Special Volume on the Integration of Language and Vision Processing, Mc Kevitt, Paul (Ed.),*, Vol. 8, No. 3 (forthcoming).

Jackson, Stuart and Noel Sharkey (1994) Grounding computational engines. In *Artificial Intelligence Review, Special Volume on the Integration of Language and Vision Processing, Mc Kevitt, Paul (Ed.)*, Vol. 8, No. 3 (forthcoming).

Joyce, James (1922) *Ulysses.* London: Faber and Faber.

Joyce, James (1939) *Finnegans Wake.* London: Faber and Faber.

Katz, J. (1972) *Semantic theory.* New York: Harper & Row.

Kosslyn, S.M. and J.R. Pomerantz (1977) Imagery, propositions and the form of internal representations. In *Cognitive Psychology*, 9, 52-76.

Kuhn T.S. (1962) *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Langton C. (1989) Artificial Life. In *Artificial Life*, C.G. Langton (Ed.), 1-47. Wokingham, UK: Addison-Wesley.

Lenat, Doug and R.V. Guha (1989) *Building large knowledge-based systems: representation and inference in the Cyc project.* Addison-Wesley.

Maaβ, Wolfgang (1994) From vision to multimodal communication: incremental route descriptions. In *Artificial Intelligence Review, Special Volume on the Integration of Language and Vision Processing, Mc Kevitt, Paul (Ed.)*, Vol. 8, No. 3 (forthcoming).

Maaβ, Wolfgang, Peter Wizinski, Gerd Herzog (1993) *VITRA GUIDE: Multimodal route descriptions for computer assisted vehicle navigation.* Bereich Nr. 93, Universitat des Saarlandes, FB 14 Informatik IV, Im Stadtwald 15, D-6600, Saarbrucken 11, Germany, EU, February.

Maybury, Mark (1991) Planning multimedia explanations using communicative acts. In *Proceedings of the Ninth American National Conference on Artificial Intelligence (AAAI-91)*, Anaheim, CA, July 14-19.

Maybury, Mark (Ed.) (1992) *Intelligent multimedia interfaces.* Menlo Park, CA: AAAI Press.

Mayhew, John and John Frisby (1981) Phychophysical and computational studies towards a theory of human stereopsis. In *Artificial Intelligence*, 17, 349-385.

MacLennan B. (1990) *Field Computation: a theoretical framework for massively parallel analog computation; parts I to IV.* Report CS-90-100, Knoxville, TN: University of Tennessee Computer Science Department.

Marr D. (1982) *Vision.* San Francisco: Freeman.

Marconi, Diego (1994) On the referential competence of some machines. In *Artificial Intelligence Review, Special Volume on the Integration of Language and Vision Processing, Mc Kevitt, Paul (Ed.)*, Vol. 8, No. 3 (forthcoming).

Meini, Cristina and Alfredo Paternoster (1994) Understanding language through vision. In *Artificial Intelligence Review, Special Volume on the Integration of Language and Vision Processing, Mc Kevitt, Paul (Ed.)*, Vol. 8, No. 1 (forthcoming).

Mc Kevitt, Paul (1986) *Building embedded representations of queries about UNIX.* Memorandum in Computer and Cognitive Science, MCCS-86-72, Computing Research Laboratory, Dept. 3CRL, Box 30001, New Mexico State University, Las Cruces, NM 88003-0001, USA.

Mc Kevitt, Paul (1991a) *Principles and practice in an operating system consultant.* Artificial Intelligence and Software Engineering, Vol. 1, Chapter on 'AI Mechanisms and techniques in practical software'. Derek Partridge (Ed.), 439-462 New York: Ablex Publishing Corporation

Mc Kevitt, Paul (1991b) *Analysing coherence of intention in natural-language dialogue.* Ph.D. Thesis, Department of Computer Science, University of Exeter, Exeter, England, EU, September.

Mc Kevitt, Paul (1992) Editorial: Natural language processing. In *Artificial Intelligence Review*, Vol. 6, No. 4, 327-332, Dec..

Mc Kevitt, P. (Ed.) (1994a) Proceedings of the Workshop on integration of natural language and vision processing. In *Twelfth American National Conference on Artificial Intelligence (AAAI-94)*, Seattle, Washington, USA, August.

Mc Kevitt, P. (Ed.) (1994b) Proceedings of the Workshop on the integration of natural language and speech processing. In *Twelfth American National Conference on Artificial Intelligence (AAAI-94)*, Seattle, Washington, USA, August.

Mc Kevitt, P. (1994c) Editorial: Visions for language. In *Artificial Intelligence Review*, Vol. 8, No. 1 (forthcoming).

Mc Kevitt, P. (1994d) Editorial: Computational processing of language and vision. In *Artificial Intelligence Review*, Vol. 8, No. 2 (forthcoming).

Mc Kevitt, P. (1994e) Editorial: Languages and visions of intelligent multimedia. In *Artificial Intelligence Review*, Vol. 8, No. 3 (forthcoming).

Mc Kevitt, Paul, Derek Partridge and Yorick Wilks (1992a) Approaches to natural language discourse processing. In *Artificial Intelligence Review*, Vol. 6, No. 4, 333-364, Dec..

Mc Kevitt, Paul, Derek Partridge and Yorick Wilks (1992b) *A survey of approaches to natural language discourse processing.* Technical Report 235, Department of Computer Science, University of Exeter, GB-EX4 4PT, Exeter, England, EU.

Mc Kevitt, Paul, Derek Partridge and Yorick Wilks (1992c) *Analysing coherence of intention in natural language dialogue.* Technical Report 227, Department of Computer Science, University of Exeter, GB-EX4 4PT, Exeter, England, EU.

Mc Kevitt, Paul, Derek Partridge and Yorick Wilks (1992d) *Why machines should analyse intention in natural language dialogue.* Technical Report 233, Department of Computer Science, University of Exeter, GB- EX4 4PT, Exeter, England, EU.

Mc Kevitt, Paul, Derek Partridge and Yorick Wilks (1992e) *Experimenting with intention in natural language dialogue.* Technical Report 234, Department of Computer Science, University of Exeter, GB- EX4 4PT, Exeter, England, EU.

Mc Kevitt, Paul and John Gammack (1994) A sensitive interface. In *Artificial Intelligence Review, Special Volume on the Integration of Language and Vision Processing*, Vol. 8, No. 3 (forthcoming).

Mc Kevitt, Paul and Cheng-ming Guo (1994) From Chinese Rooms to Irish Rooms: new words for visions of language. In *International Workshop on Dir-*

ections of Lexical Research, August, 1994, Beijing, China.

Morgan, Simon and Paul Mc Kevitt (1994a) Algebraic structure as vision for language. In *Artificial Intelligence Review, Special Volume on the Integration of Language and Vision Processing*, Vol. 8, No. 1 (forthcoming).

Morgan, Simon and Paul Mc Kevitt (1994b) Other way round. In *Artificial Intelligence Review, Special Volume on the Integration of Language and Vision Processing*, Vol. 8, No. 3 (forthcoming).

Narayanan, A., Ford, L. Manuel, D., Tallis, D. and Yazdani, M. (1994) Language animation. In *Proceedings of the Workshop on integration of natural language and vision processing, Mc Kevitt, Paul (Ed.), Twelfth American National Conference on Artificial Intelligence (AAAI-94)*, Seattle, Washington, USA, August.

Neumann, B. and H.-J. Novak (1986) NAOS: Ein System zur naturalichsprachlichen Beschreibung zeitveranderlicher Szenen. In *Informatik. Forschung und Entwicklung*, 1(1): 83-92.

Partridge, Derek (1991) *A new guide to Artificial Intelligence.* Norwood, New Jersey: Ablex Publishing Corporation.

Popper K.R. (1972) *Objective Knowledge.* Oxford: Clarendon Press.

Procter, P. (1978) *Longman Dictionary of Contemporary English.* London: Longman.

Pylyshyn, Zenon (1973) What the mind's eye tells the mind's brain: a critique of mental imagery. In *Psychological Bulletin*, 80, 1-24.

Retz-Schmidt, Gudala (1991) Recognizing intentions, interactions, and causes of plan failures. In *User Modelling and User-Adapted Interaction*, 1: 173-202.

Retz-Schmidt, Gudala and Markus Tetzlaff (1991) *Methods for the intentional description of image sequences.* Bereich Nr. 80, Universitat des Saarlandes, FB 14 Informatik IV, Im Stadtwald 15, D-6600, Saarbrucken 11, Germany, EU, August.

Rich, Elaine and Kevin Knight (1991) *Artificial Intelligence.* New York: McGraw-Hill.

Rowe, Jon and Paul Mc Kevitt (1991) An emergent computation approach to natural language processing. In *Proceedings of the Fourth Irish Conference on Artificial Intelligence and Cognitive Science*, University College Cork, IRL- Cork, Ireland, European Union (EU), September.

Schank, Roger C. (1972) Conceptual dependency: a theory of natural language understanding. In *Cognitive Psychology*, 3(4): 552-631.

Schank, Roger C. (1973) Identification and conceptualizations underlying natural language. In *Computer Models of Thought and Language*, R. Schank and K. Kolby (Eds.). San Francisco, CA: Wh Freeman and Co..

Schank, Roger C. (1975) *Conceptual information processing.* Fundamental Studies in Computer Science, 3. Amsterdam: North-Holland.

Schank, Roger and Andrew Fano (1994) Memory and expectations in learning, language and visual understanding. In *Artificial Intelligence Review, Special Volume on the Integration of Language and Vision Processing, Mc Kevitt, Paul (Ed.),*, Vol. 8, No. 1 (forthcoming).

Searle, J.R. (1980) Minds, brains and programs. In *Behaviour and Brain Sciences*, 3: 417-424.

Searle, J.R. (1984) *Minds, brains and science.* London: Penguin Books.

Searle, J.R. (1990) Is the brain's mind a computer program?. In *Scientific American*, 262: 26-31.

Sinclair, John (Ed.) (1987) *Looking Up: an account of the COBUILD project in lexical computing.* London: Collins.

Stannett M. (1990) X-machines and the Halting Problem: Building a Super-Turing Machine. In *J. Formal Aspects of Computing 2*, pp 331-41.

Stock, Oliviero (1991) Natural language and exploration of an information space: the ALFresco Interactive system. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI-91)*, 972-978, Darling Harbour, Sydney, Australia, August.

Tao, Li (1989) *How to read faces.* London: Hamlyn.

Wahlster, Wolfgang (1988) *One word says more than a thousand pictures: On the automatic verbalization of the results of image sequence analysis.* Bereich Nr. 25, Universitat des Saarlandes, FB 14 Informatik IV, Im Stadtwald 15, D-6600, Saarbrucken 11, Germany, EU, February.

Wahlster, Wolfgang, Elisabeth André, Wolfgang Finkler, Hans-Jurgen Profitlich, and Thomas Rist (1992) Plan-based integration of natural language and graphics generation. In *Artificial Intelligence, Special Issue On Natural Language Generation*, 63: 387-427.

Waltz, David (1975) Understanding line drawings of scenes with shadows. In *The psychology of computer vision*, Winston, P.H. (Ed.), 19-91. New York: McGraw-Hill.

Wilks, Yorick (1973) An artificial intelligence approach to machine translation. In *Computer Models of Thought and Language*, R. Schank and K. Kolby (Eds.). San Francisco, CA: Wh Freeman and Co..

Wilks, Yorick (1975a) Preference semantics. In *Formal semantics of natural language*, Keenan, Edward (Ed.). Cambridge, United Kingdom: Cambridge University Press. Also as, Memo AIM-206, Artificial Intelligence Laboratory, Stanford University, Stanford, California, USA, July 1973.

Wilks, Yorick (1975b) An intelligent analyzer and understander of English. In *Communications of the ACM*, 18(5), 264-274, May. Also in, *Readings in Natural Language Processing*, Barbara Grosz, Karen Sparck Jones and Bonnie Webber (Eds.), 193-204, Los Altos, California: Morgan Kaufmann, 1986.

Wilks, Yorick (1975c) A preferential, pattern-seeking semantics for natural language inference. In *Artificial Intelligence*, 6, 53-74.

Wilks, Y. (1977) Good and bad arguments about semantic primitives. In *Communication and cognition*, Vol. 10 No. 3/4, 181-221.

Wilks, Yorick (1994) Language, vision and metaphor. In *Artificial Intelligence Review, Special Volume on the Integration of Language and Vision Processing, Mc Kevitt, Paul (Ed.)*, Vol. 8, No. 1 (forthcoming).

Wilks, Yorick and Afzal Ballim (1987) Multiple agents and the heuristic ascription of belief. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI-87)*, 119-124, Milan, Italy, August.

Wilks, Y., and Okada, N., (Eds.) (in press) *Computer Language & Vision Across the Pacific.* Norwood, NJ: Ablex.

Wittgenstein, Ludwig (1963) *Philosophical Investigations (translated by G.E. Anscombe).* Oxford: Blackwell.

Wright, Ron, E. and Young, D.A. (1990) The cognitive modalities (CM) system of knowledge representation and its potential application in C31 systems. In *Proceedings of The 1990 Symposium on Command and Control Research SAIC Report SAIC-90/1508*, pp. 373-381.

Young, D.A. (1983) Interactive modal meaning as the mental basis of syntax. In *Medical Hypotheses*, 10:5.